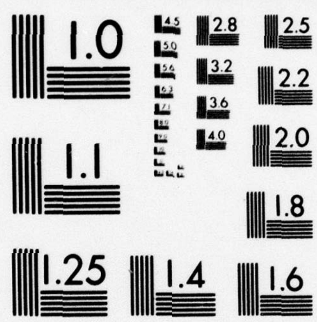
AD-A063 887 OHIO STATE UNIV COLUMBUS DEPT OF GEODETIC SCIENCE F/G 8/5
THE OPERATIONAL APPROACH TO PHYSICAL GEODESY.(U)
OCT 78 H MORITZ F19628-76-C-0010
UNCLASSIFIED DGS-277 AFGL-TR-78-0281 NL

| OF |
AD
A063 887

END
DATE
FILMED
3-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFGL-TR-78-0281

LEVEL II

THE OPERATIONAL APPROACH TO PHYSICAL GEODESY

Helmut Moritz

The Ohio State University
Research Foundation
Columbus, Ohio 43212

Scientific Report No. 21
October, 1978

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (19) (18) AFGL-TR-78-0281	2. GOVT ACCESSION NO. (9)	3. RECIPIENT'S CATALOG NUMBER Interim rept.
4. TITLE (and Subtitle) (6) THE OPERATIONAL APPROACH TO PHYSICAL GEODESY.		5. TYPE OF REPORT & PERIOD COVERED Scientific. Interim. Scientific Report No. 21
		6. PERFORMING ORG. REPORT NUMBER Report No. 277
7. AUTHOR(s) (10) HELMUT MORITZ		8. CONTRACT OR GRANT NUMBER(s) (15) F19628-76-C-0010
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Geodetic Science The Ohio State University Columbus, Ohio 43210		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62101 (16) 760003AG (17) 03
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory Hanscom AFB, Massachusetts 01731 Contract Monitor - Bela Szabo/LW		12. REPORT DATE (11) October 1978
		13. NUMBER OF PAGES 66 pages
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) (12) 67p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

(14) DGS-277, SCIENTIFIC-21

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

TECH, OTHER

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Geodesy, Gravity, Least-squares Methods, Improperly Posed Problems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The operational approach to physical geodesy starts from the measurements and asks how they can be used in the best way to determine the earth's figure and gravitational field.

All measurements can be represented as nonlinear functionals of the potential and of systematic parameters such as station coordinates. After linearization we obtain an improperly posed problem, to which three standard → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

400 254 Unclassified

79 01 26 019

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

cont.

methods of solving such problems are applied: a change in the solution space, a variational principle according to Tichonov, and a statistical approach. All three approaches seem to converge on collocation with kernel functions and least-squares collocation. Some alternatives are also ~~di~~ discussed.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

FOREWORD

This report was prepared by Dr. Helmut Moritz, Professor, Technische Hochschule in Graz, and Adjunct Professor, Department of Geodetic Science of The Ohio State University, under Air Force Contract No. F19628-76-C-0010, The Ohio State University Research Foundation Project No. 710334, Project Supervisor, Urho A. Uotila, Professor, Department of Geodetic Science. The contract covering this research is administered by the Air Force Geophysical Laboratories, L.C. Hanscom Air Force Base, Bedford, Massachusetts, with Mr. Bela Szabo/LW, Project Scientist.

CONTENTS

1. The Operational Approach

There are essentially two possible approaches to physical geodesy (as also to other natural sciences): they might be called the model approach and the operational approach. Essentially, the first approach starts from a theory, the second from the observations. Obviously, the two approaches are closely related to the deductive method and the inductive method in the natural sciences.

In the model approach, one starts from a mathematical model or from a theory and then tries to fit this model to reality, for instance by determining the parameters of this model from observations. The classical geodetic example are the centuries-old attempts to determine the parameters of an earth ellipsoid by observation, from the old grade measurements to modern satellite observations.

Perhaps the most elaborate form of this model approach is the boundary-value problem of physical geodesy in the formulation of Molodensky. It has a mathematically enormously interesting and deep theory and is practically highly significant, as the many gravimetric geoid determinations and computations of deflections of the vertical show. However, this approach has its weeknesses: the required continuous gravity coverage is practically not realizable; on the other hand, many other important data cannot be incorporated into this theory. The model selects its data.

At present we have a great number of geodetic measurements of very different types, from terrestrial angle and distance measurements to satellite data of various kinds. The question arises: how can we use and combine all these data in the best possible way. This is the operational approach.

Let us summarize. In the model approach one asks: how can I best determine my model by suitable observations? In the operational approach one asks: how can I make best use of all my observations?

As a matter of fact, the two approaches do not compete with each other; each one incorporates important aspects, and the two approaches mutually complement each other.

The operational approach to physical geodesy has come up at a relatively recent date, when a huge number of measurements of new types was available and when it turned out that the classical, especially the gravimetric approach failed to give a complete answer in view of the lack in gravity data.

In geometrical geodesy already least-squares adjustment is in the spirit of an operational approach (how can I best use all my measurements). In physical geodesy, operational methods have been known under the names "least-squares collocation", "integrated geodesy", "operational geodesy". All these methods are very similar; they all aim at an adequate treatment of the gravity field, in addition to an adjustment of measuring errors. They all use quadratic minimum principles incorporating not only the measuring errors, but also the anomalous gravity field.

The purpose of the present report is a systematic treatment of the operational approach. Obviously, the determination of the gravity field, of station coordinates, etc., from discrete measurements does not in general admit a unique solution; it is an "improperly posed problem". The question is: Which practically feasible possibilities exist? Is there a genuine alternative to least-squares collocation and related methods, which is, perhaps, even simpler?

This question is treated by using modern techniques for solving improperly posed problems. They all seem to converge on collocation by means of kernel functions and least-squares collocation; but we shall also try to point out alternatives.

2. Measurements as Nonlinear Functionals

Every geodetic measurement depends:

1. on one or several points in space;
2. on the earth's gravitational field.

Symbolically we may write:

$$l = F(X,V) . \qquad (2\text{-}1)$$

Here l denotes the measurement under consideration, V denotes the gravitational potential, and the vector X comprises the coordinates of the points to which the measurement refers, and possibly other parameters. For instance, if we have two points P and Q and if we use rectangular coordinates xyz referred to some Cartesian reference system, then

$$X = \left[x_P, y_P, z_P, x_Q, y_Q, z_Q\right]^T , \qquad (2\text{-}2)$$

T denoting the transpose (in the sequel, vectors will be column vectors unless the contrary is stated).

The symbol F denotes any functional dependence on X and V. With respect to X, it is an ordinary function; but it is not necessarily an ordinary function of V but may involve first and higher derivatives of V, integrals, etc. In the terminology of functional analysis, F is a (nonlinear) functional of X and V.

Denote the number of components of x by m; then X may be said to belong to m-dimensional Euclidean space R_m. The function V may be considered to belong to some set, or space, H of harmonic functions. Then, in the jargon of modern mathematics, the functional F is a mapping of the product space $R_m \times H$ into R, the real number line:

$$F: R_m \times H \to R , \qquad (2\text{-}3)$$

which, in the language of plain mortals, means simply that F associates, to each harmonic function from the set H and to each vector R_m, a real number which represents the numerical value of the observation l.

More intuitively we may say that F is nothing else but a prescription for computing a number l from a given vector X and a given function V: if X and V are supposed to be known, then it must be possible to find, in an unambiguous way, the value of l. In other terms, F denotes an operation to be performed on X and V, the result of which is a real number.

Thus, functionals are special cases of operators: for an operator, the result of the operation may be quite general: a vector or another function; if the result is a real number, the operator is called a functional. Best known from functional analysis (cf. Kantorovich and Akilov, 1964; Meissl, 1975; Tscherning, 1978) are linear operators and linear functionals. Nonlinear operations are basic in the so-called "Modern Analysis" (Dieudonné, 1960; Loomis and Sternberg, 1968).

Our functionals (2-1) will, in general, be nonlinear; in the next section, we shall describe how they can be linearized.

The physical meaning of such functionals F will be clear from the examples given below.

Instead of the gravitational potential V, we may also use the gravity potential W, defined in the usual way by

$$W = V + \frac{1}{2}\omega^2(x^2 + y^2) , \qquad (2\text{-}4)$$

ω denoting the angular velocity of the earth's rotation and the z-axis coinciding with the earth's axis of rotation. For the present purpose it is appropriate to regard ω as constant and the rotation axis as invariable with respect to the earth's body: the very small deviations from this idealized situation can easily be taken into account by corrections, without changing the essential picture.

Throughout this report, the system xyz will thus be defined as follows: the origin is at the earth's center of mass, the z-axis coincides with the (mean) rotation axis, and the x-axis goes through the (mean) Greenwich meridian.

Then, instead of (2-1), we have

$$l = F(X,W). \tag{2-5}$$

As W is expressed in terms of V and X by (2-4), the relations (2-1) and (2-5) are equivalent; as a matter of fact, the letter F denotes different functionals in each of the two cases.

Let us now illustrate these abstract considerations by means of concrete examples.

Astronomical and gravimetric observations. The gravity vector G is expressed in terms of gravity g and astronomical latitude Φ and longitude Λ by means of the well-known relation

$$-G = \begin{bmatrix} g\cos\Phi\cos\Lambda \\ g\cos\Phi\sin\Lambda \\ g\sin\Phi \end{bmatrix} \tag{2-6}$$

(cf. Heiskanen and Moritz, 1967, p.57). On the other hand, G is the gradient of the gravity potential W :

$$G = \operatorname{grad} W = \begin{bmatrix} W_x \\ W_y \\ W_z \end{bmatrix}, \tag{2-7}$$

W_x denoting the partial derivative

$$W_x = \frac{\partial W}{\partial x} \tag{2-8}$$

and similarly for W_y and W_z. Comparing (2-6) and (2-7) and solving for Φ, Λ, and g we obtain

$$\Phi = \tan^{-1} \frac{-W_z}{\sqrt{W_x^2 + W_y^2}}, \tag{2-9}$$

$$\Lambda = \tan^{-1} \frac{W_y}{W_x} \quad , \tag{2-10}$$

$$g = \sqrt{W_x^2 + W_y^2 + W_z^2} \quad . \tag{2-11}$$

These equations have the form (2-5): they express the observables Φ, Λ, g in terms of the potential W, not as ordinary functions of W, but as nonlinear functionals involving the operation of differentiation. Let X denote the coordinate vector of the observation station:

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad . \tag{2-12}$$

Then, as W_x, W_y, W_z are functions of x, y, z, the expressions (2-9) to (2-11) do, in fact, also depend on X, in agreement with (2-5).

Angle and distance measurements. The observables: azimuth A, zenith distance Z, and distance S between two points P and Q, can be expressed in terms of the coordinate differences

$$\begin{aligned} \Delta x &= x_Q - x_P \quad , \\ \Delta y &= y_Q - y_P \quad , \\ \Delta z &= z_Q - z_P \quad , \end{aligned} \tag{2-13}$$

as follows (Heiskanen and Moritz, 1967, p.219):

$$A = \tan^{-1} \frac{-\Delta x \sin\Lambda + \Delta y \cos\Lambda}{-\Delta x \sin\Phi \cos\Lambda - \Delta y \sin\Phi \sin\Lambda + \Delta z \cos\Phi} \quad , \tag{2-14}$$

$$Z = \cos^{-1} \frac{\Delta x \cos\Phi \cos\Lambda + \Delta y \cos\Phi \sin\Lambda + \Delta z \sin\Phi}{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}} \quad , \tag{2-15}$$

$$s = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \quad . \tag{2-16}$$

Again, these equations have the form (2-5); the vector X is now

$$X = \left[x_P,\ y_P,\ z_P,\ x_Q,\ y_Q,\ z_Q\right]^T, \tag{2-17}$$

comprising the coordinates of both points P and Q, and the dependence on the potential W is implicitly through Φ and Λ as expressed by (2-9) and (2-10); hence A and Z are, in fact, nonlinear functionals of W. Note that these observables depend on the target point Q only because its coordinates enter into Δx, Δy, Δz; on the observation station P they depend in the same way, but there is an additional dependence on P because Φ and Λ, and hence W_x, W_y, W_z, refer to this point.

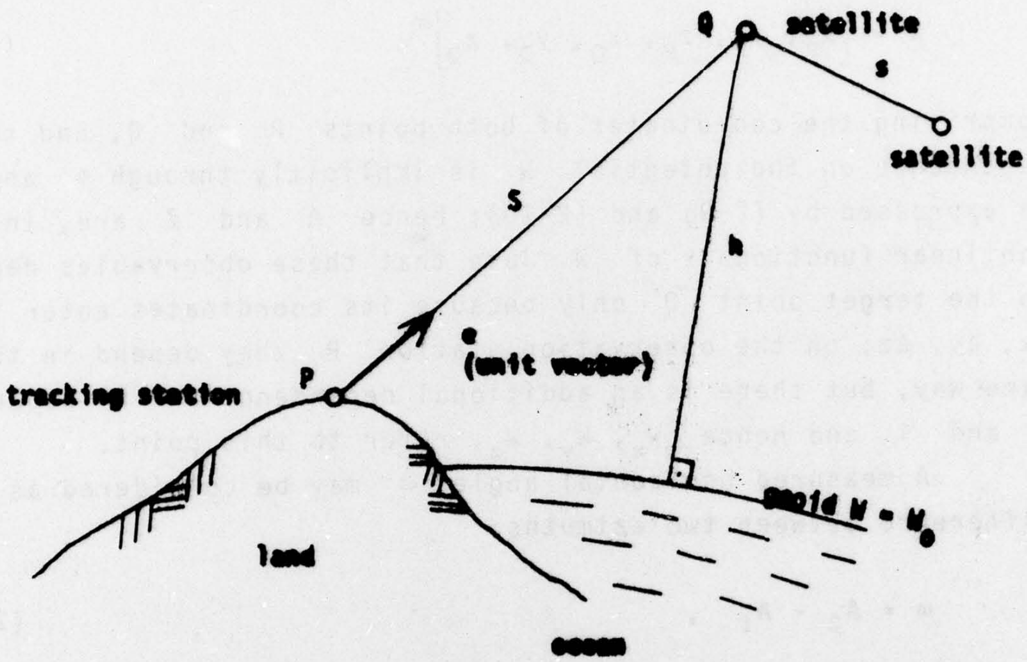
A measured horizontal angle ω may be considered as the difference between two azimuths:

$$\omega = A_2 - A_1 \ , \tag{2-18}$$

measured at an observation station P to two targets Q_1 and Q_2. Both azimuths A_1 and A_2 may be expressed by (2-14); the resulting expression for ω clearly involves the coordinates of P, Q_1, and Q_2, so that, in the present case, the vector X consists of 9 components, which are the coordinates of these three points; we again get a nonlinear functional of form (2-5).

It goes without saying that the functional expression (2-16) for the distance (2-16) is also a special case of (2-5), in which there simply is no factual dependence on W: measured straight distances between two fixed points do not depend on the gravity field.

<u>Satellite observations</u>. Consider a <u>distance</u> S measured from a ground station P to a satellite Q by laser or radar. (Cf. Fig. 2-1; for a non-technical and compact review of various techniques see (Cordova, 1977).)



e	direction observation
S	distance measurement
dS/dt	doppler observation
h	satellite altimetry
ds/dt	satellite-to-satellite tracking (doppler)

Figure 2-1. Satellite techniques

Such a distance can again be represented by (2-16) but, if we operate in the orbital mode, the coordinates of Q can be further expressed by the six orbital elements $p_1, p_2, \ldots, p_6$ of some reference orbit and the coefficients J_{nm} and K_{nm} of the expansion of the earth's gravitational potential V in terms of spherical harmonics. Thus S will have the form of some function

$$S = S(x_P, y_P, z_P; p_1, p_2, \ldots, p_6; J_{nm}, K_{nm}) \; . \qquad (2\text{-}19)$$

This is a functional of form (2-1). The vector X is given now by

$$X = \left[x_P, y_P, z_P; p_1, p_2, \ldots, p_6 \right]^T ; \qquad (2\text{-}20)$$

it comprises station coordinates and orbital parameters. The spherical-harmonic coefficients J_{nm} and K_{nm} may be expressed in terms of V by well-known integral formulas (of type of eq.(1-76) of (Heiskanen and Moritz, 1967, p.31)), which explains the functional dependence on V.

The change of distance S with respect to time t, that is, the range-rate dS/dt, can be measured by doppler observations. By integrating dS/dt with respect to t from t_1 to t_2, one obtains distance differences $S_2 - S_1$. By photographing the satellite against the background of stars one finds the right ascension and the declination of the spatial direction PQ, or in other terms, the unit vector e of this direction (Fig. 2-1). All these observables have the same mathematical structure as (2-19): they are again functionals of type (2-1), the vector X being given by (2-20).

Satellite altimetry can be considered to measure the height h of a satellite above the geoid: the ocean surface reflects a radar signal emitted by the satellite, and under idealized conditions, this surface coincides with the geoid. We claim that h can again be expressed as a functional of type (2-1), with

$$X = \left[p_1, p_2, \ldots, p_6\right]. \tag{2-21}$$

This is true if it is possible, given X and the potential function $V(x,y,z)$, to compute h. In fact, assume the gravitational potential $V(x,y,z)$ to be known as a function of position at all points outside and on the earth's surface. Then the gravity potential function W is also known by (2-4), and consequently the geoid is an equipotential surface

$$W(x,y,z) = W_o = \text{const.} \tag{2-22}$$

Now, the satellite orbit can be computed from the parameters p_k of the reference orbit and the gravitational potential V, and the position Q of the satellite along the orbit is uniquely determined by giving the corresponding instant t (which we assume to be known). Thus both the geoid and the satellite position Q are determined, and so is h, as the length of the perpendicular from Q to the geoid. Therefore, by the definition of the functional (2-1) given above, the satellite altimeter measurement h is, in fact, such a functional.

The data of satellite-to-satellite tracking are time changes of the distance s between two satellites (Fig. 2-1). Such a range rate ds/dt is again measured by means of the doppler principle. At present one generally uses one high and one low satellite, but the use of two low satellites which are close to each other is also possible. Considerations analogous to the preceding ones make it obvious that ds/dt has again the form (2-1), the vector X comprising now the 6+6 elements of the two reference orbits.

Satellite gradiometry is designed to measure elements (or linear combinations of elements) of the second-order gradient tensor

$$\begin{bmatrix} V_{xx} & V_{xy} & V_{xz} \\ V_{xy} & V_{yy} & V_{yz} \\ V_{xz} & V_{yz} & V_{zz} \end{bmatrix} , \qquad (2\text{-}23)$$

which is a symmetric matrix formed by the second derivatives of the potential V with respect to the coordinates xyz. Any second-order gradient, say V_{xz}, depends on position:

$$V_{xz} = V_{xz}(x,y,z) \; . \qquad (2\text{-}24)$$

It has, therefore, the form (2-1), with

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} ; \qquad (2\text{-}25)$$

the prescription for computing the functional F in the present case consists in differentiating V with respect to x and z and taking V_{xz} at the point with coordinates (2-25).

<u>Very-long-baseline interferometry</u> measures the delay τ, with which a radio signal emitted from an extragalactic radio source is received at two different places P and Q (Macdoran, 1973). By multiplying τ with the light velocity c we get the projection

$$D = \vec{PQ}.e \qquad (2\text{-}26)$$

of the vector $\vec{PQ}$ connecting the two points onto the direction (supposed known) to the radio source represented by the unit vector e. Similarly to a distance measurement (2-16), D does not depend on the gravity field, and we have a special case of (2-1) with X being given by (2-17) and with no explicit dependence on V.

These examples should make it obvious that <u>all</u> geodetic measurements, without exception, can be represented as functionals (2-1) or (2-5). This simple and general fact will be basic for the considerations to follow.

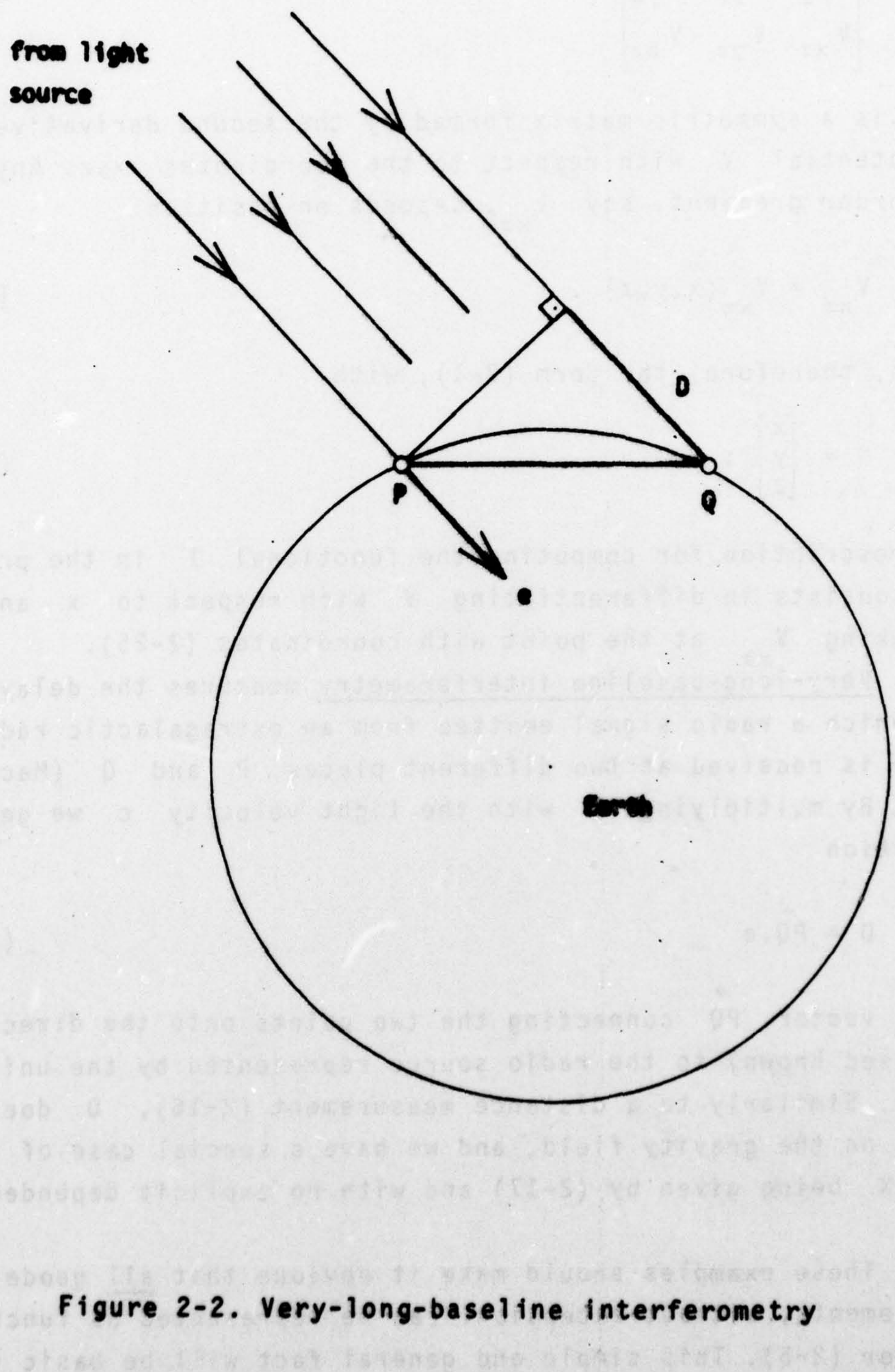


Figure 2-2. Very-long-baseline interferometry

It is clear that we have taken into account only the geometrical and gravitational structure of the problem. We have abstracted from random and systematic errors, nongravitational effects, etc. Random errors will be considered later in this report, and systematic effects are assumed to have been removed by appropriate corrections. If necessary, systematic parameters can be included in the vector X in (2-1) or (2-5).

3. Linearization

Every observation l gives an equation of type (2-1) or (2-5). We thus obtain a system of functional equations

$$\begin{aligned} l_1 &= F_1(X,W) \; , \\ l_2 &= F_2(X,W) \; , \\ &\vdots \\ l_q &= F_q(X,W) \; , \end{aligned} \tag{3-1}$$

which are to be solved for the unknown parameters X and the potential function W.

Since the functionals $F_1, F_2, \ldots, F_q$ are non-linear, the system (3-1) is very difficult to handle directly. The usual procedure with difficult nonlinear problems is to linearize them by Taylor's theorem.

Let us introduce an approximate value X_o for the vector X and an approximation U to the gravity potential W. The function U is called the normal potential; it is generally taken to be the external gravity potential of an equipotential ellipsoid (cf. Heiskanen and Moritz, 1967, sec. 2-7).

We put

$$X = X_o + \delta X \; , \tag{3-2}$$

$$W = U + T \; , \tag{3-3}$$

where the differences $\delta X = X - X_o$ and $T = W - U$ are considered to be small; T is called the <u>anomalous potential</u> (<u>ibid</u>., sec.2-13).

Thus (2-5) becomes

$$l = F(X_o + \delta X,\ U+T) \tag{3-4}$$

and a Taylor expansion gives

$$l = F(X_o, U) + a^T \delta X + LT \tag{3-5}$$

plus higher order terms, which we neglect. Here a is the column vector of ordinary partial derivatives

$$a_k = \frac{\partial F}{\partial X_k}(X_o, U) \tag{3-6}$$

of F with respect to the component X_k of the parameter vector X, taken for the approximate values X_o and U; a^T is the corresponding row vector, so that $a^T \delta X$ is a scalar product. The term LT is less elementary: it expresses a linear operator L acting on the function T. The meaning of this will be clear from the examples to follow.

By means of the substitution

$$\delta l = F(X, W) - F(X_o, U) \ , \tag{3-7}$$

the nonlinear system (3-1) thus becomes the linear system

$$\begin{aligned} \delta l_1 &= a_1^T \delta X + L_1 T \ , \\ \delta l_2 &= a_2^T \delta X + L_2 T \ , \\ &\ \vdots \\ \delta l_q &= a_q^T \delta X + L_q T \ . \end{aligned} \tag{3-8}$$

The linearization process will be made clear by considering some basic special cases.

<u>Astronomical and gravimetric observations.</u> The equations to be linearized are (2-9), (2-10), and (2-11). Here we are considerably helped by the fact that these expressions are just ordinary functions of W_x, W_y, W_z; X is simply the coordinate vector (2-12).

Therefore, we first linearize the gradient vector (2-7). Using index notation, we write $x = x_1$, $y = x_2$, $z = x_3$ and

$$\text{grad } W = \begin{bmatrix} W_x \\ W_y \\ W_z \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} . \tag{3-9}$$

Thus

$$W_i = \frac{\partial W}{\partial x_i} \qquad (i = 1,2,3) . \tag{3-10}$$

The derivatives are taken at the original point with coordinates

$$X = \left[x_k\right] \qquad (k = 1,2,3) . \tag{3-11}$$

The approximation point is

$$X_o = \left[x_k^o\right] ; \tag{3-12}$$

in obvious notation,

$$x_k = x_k^o + \delta x_k . \tag{3-13}$$

Then (3-10) becomes

$$W_i = \frac{\partial W}{\partial x_i} = \left(\frac{\partial W}{\partial x_i}\right)_o + \left(\frac{\partial^2 W}{\partial x_i \partial x_j}\right)_o \delta x_j , \tag{3-14}$$

using the summation convention (summation over the repeated index j). The notation $(\ )_o$ indicates that the respective quantity is to be taken at the approximation point (3-12); W_i is, of course, taken at the original point X.

We now introduce $W = U + T$ and obtain

$$\frac{\partial W}{\partial x_i} = \left(\frac{\partial U}{\partial x_i}\right)_o + \left(\frac{\partial T}{\partial x_i}\right)_o + \left(\frac{\partial^2 U}{\partial x_i \partial x_j}\right)_o \delta x_j + \left(\frac{\partial^2 T}{\partial x_i \partial x_j}\right)_o \delta x_j \ . \qquad (3\text{-}15)$$

The last term is already of second order (T and δx_j are first-order quantities) and will be neglected. We further put

$$\left(\frac{\partial U}{\partial x_i}\right)_o = U_i \ , \qquad (3\text{-}16)$$

$$\left(\frac{\partial T}{\partial x_i}\right)_o = T_i \ , \qquad (3\text{-}17)$$

$$\left(\frac{\partial^2 U}{\partial x_i \partial x_j}\right)_o = M_{ij} \ . \qquad (3\text{-}18)$$

Thus (3-15) becomes finally

$$W_i = U_i + T_i + M_{ij}\delta x_j \ , \qquad (3\text{-}19)$$

completing the linearization of the gravity vector (3-9).

The straightforward way to linearize equations (2-9) to (2-11) is to substitute (3-19) into these equations and to expand the functions in the usual way by Taylor's theorem, considering the fact that the second and third term on the right-hand side of (3-19) are small. This is simple but laborious; more efficient is an indirect procedure.

We combine (2-6) and (2-7) into the equation

$$\begin{aligned} W_1 &= -g\cos\phi\cos\Lambda \ , \\ W_2 &= -g\cos\phi\sin\Lambda \ , \\ W_3 &= -g\sin\phi \ ; \end{aligned} \qquad (3\text{-}20)$$

here all quantities refer to the original point X.

In an analogous way we write

$$\begin{aligned} U_1 &= -\gamma\cos\phi\cos\lambda \quad , \\ U_2 &= -\gamma\cos\phi\sin\lambda \quad , \\ U_3 &= -\gamma\sin\phi \quad . \end{aligned} \tag{3-21}$$

Here all quantities refer to the approximation point: γ is normal gravity, and ϕ and λ are normal latitude and longitude. Cf. (Heiskanen and Moritz, 1967, p.315), where these normal geographical coordinates have been denoted by ϕ^* and λ^*.

We put

$$\begin{aligned} \Phi &= \phi + \delta\Phi \ , \\ \Lambda &= \lambda + \delta\Lambda \ , \\ g &= \gamma + \delta g \ , \end{aligned} \tag{3-22}$$

substitute into (3-20) and expand by Taylor. The result is readily found to be

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} + Q \begin{bmatrix} \gamma\delta\Phi \\ \gamma\cos\phi\delta\Lambda \\ \delta g \end{bmatrix} \ , \tag{3-23}$$

where the matrix

$$Q = \begin{bmatrix} \sin\phi\cos\lambda & \sin\lambda & -\cos\phi\cos\lambda \\ \sin\phi\sin\lambda & -\cos\lambda & -\cos\phi\sin\lambda \\ -\cos\phi & 0 & -\sin\phi \end{bmatrix} \tag{3-24}$$

is obtained by differentiation of (3-21).

On the other hand we have (3-19), which may be written in the form

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \end{bmatrix} + \operatorname{grad}T + M\delta X \ . \tag{3-25}$$

The matrix Q is easily seen to be orthogonal (why?); therefore its inverse is simply the transpose:

$$Q^{-1} = Q^T \ . \tag{3-26}$$

Therefore, the comparison of (3-23) and (3-25) gives

$$\begin{bmatrix} \gamma\delta\Phi \\ \gamma\cos\phi\delta\Lambda \\ \delta g \end{bmatrix} = Q^T M \delta X + Q^T \operatorname{grad} T \ , \tag{3-27}$$

which completes the linearization of astronomical latitude Φ and longitude Λ and of measured gravity g.

It is evident that (3-27) is, indeed, a linear function of the components δx, δy, δz of the vector δX. As regards $Q^T \operatorname{grad} T$, it gives for each difference $\delta\Phi$, $\delta\Lambda$, δg, a linear expression of the form

$$\alpha_1 \frac{\partial T}{\partial x} + \alpha_2 \frac{\partial T}{\partial y} + \alpha_3 \frac{\partial T}{\partial z} = LT \ . \tag{3-28}$$

The operation expressed by the functional L consists in forming the partial derivatives and taking a linear combination of them. Since differentiation is a linear operation, L is indeed a linear functional.

<u>Direction and distance measurements</u>. The straightforward approach is to differentiate equations (2-14), (2-15) and (2-16), as outlined in (Heiskanen and Moritz, 1967, pp.220-221). The result will be differential formulas of form of eq. (5-83), <u>ibid</u>. The actual work is, however, quite cumbersome though not difficult.

Again, an indirect approach might be preferable. We put

$$\begin{bmatrix} s\sin Z\cos A \\ s\sin Z\sin A \\ s\cos Z \end{bmatrix} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = Y \ . \tag{3-29}$$

Then the vector Y so defined is related to the difference vector

$$\Delta X = \begin{bmatrix} x_Q - x_P \\ y_Q - y_P \\ z_Q - z_P \end{bmatrix} \tag{3-30}$$

by the linear transformation

$$Y = R\Delta X \quad , \tag{3-31}$$

where R is the orthogonal matrix

$$R = \begin{bmatrix} -\sin\Phi\cos\Lambda & -\sin\Phi\sin\Lambda & \cos\Phi \\ -\sin\Lambda & \cos\Lambda & 0 \\ \cos\Phi\cos\Lambda & \cos\Phi\sin\Lambda & \sin\Phi \end{bmatrix} . \tag{3-32}$$

This is clear because u,v,w can be interpreted as rectangular coordinates in a local system in which the w-axis has the direction of the gravity vector and the axes u and v point north and east, respectively; the matrix R is formed by the components in the xyz system, of the unit vectors e', e'', n corresponding to the uvw coordinate axes (Heiskanen and Moritz, 1967, pp.218-219).

The differentiation of (3-29) gives, in analogy to (3-23)

$$\delta Y = S \begin{bmatrix} \sigma\delta Z \\ \sigma\sin\zeta\delta A \\ \delta s \end{bmatrix} \tag{3-33}$$

where S is the orthogonal matrix

$$S = \begin{bmatrix} \cos\zeta\cos\alpha & -\sin\alpha & \sin\zeta\cos\alpha \\ \cos\zeta\sin\alpha & \cos\alpha & \sin\zeta\sin\alpha \\ -\sin\zeta & 0 & \cos\zeta \end{bmatrix} . \tag{3-34}$$

Here we have designed by α,ζ,σ the "normal" equivalents of the observables A,Z,s, so that

$$\begin{aligned} A &= \alpha + \delta A , \\ Z &= \zeta + \delta Z , \\ s &= \sigma + \delta s . \end{aligned} \tag{3-35}$$

The quantities α,ζ,σ can be computed from (2-14), (2-15), and (2-16) by using approximate coordinates X and replacing Φ,Λ by ϕ,λ.

By differentiation of (3-31), on the other hand, we find

$$\delta Y = R\delta\Delta X + \delta R\Delta X \quad . \tag{3-36}$$

The combination of (3-33) and (3-36) gives, in view of the orthogonality of the matrix S,

$$\begin{bmatrix} \sigma\delta Z \\ \sigma\sin\zeta\delta A \\ \delta s \end{bmatrix} = S^T R\delta\Delta X + S^T\delta R\Delta X \ . \tag{3-37}$$

The second term on the right-hand side is easily found in an indirect way. The matrix R, by (3-32), depends on Φ and Λ; therefore δR will be a linear function of $\delta\Phi$ and $\delta\Lambda$. The term $S^T\delta R\Delta X$ represents, therefore, the effect of $\delta\Phi$ and $\delta\Lambda$ on δZ and δA (there is, evidently, no effect on δs because s is independent of the gravity field); this is nothing else but the well-known effect of the deflection of the vertical on azimuth A and zenith distance Z.

The effect on the zenith distance Z is

$$\partial Z = \xi\cos\alpha + \eta\sin\alpha \tag{3-38}$$

(Heiskanen and Moritz, 1967, p.190, eq.(5-20)) and on the azimuth

$$\partial A = \xi \sin\alpha \cot\zeta + \eta(\tan\phi - \cos\alpha\cot\zeta) \tag{3-39}$$

(<u>ibid</u>., p.186, eq.(5-13)). We have used the symbols ∂Z and ∂A to indicate the partial influence, on Z and A, of the changes $\delta\Phi$ and $\delta\Lambda$, which are related to the deflection components ξ and η by

$$\xi = \delta\Phi \ , \qquad \eta = \delta\Lambda\cos\phi \ . \tag{3-40}$$

The comparison of (3-24) and (3-32) shows that, for $\Phi = \phi$ and $\Lambda = \lambda$,

$$R = -Q^T \ ; \tag{3-41}$$

the geometrical interpretation of this fact is left to the reader.

In view of the relations (3-38) to (3-41), eq.(3-37) takes the final form

$$\begin{bmatrix} \sigma\delta Z \\ \sigma\sin\zeta\delta A \\ \delta s \end{bmatrix} = -S^T Q^T \begin{bmatrix} \delta x_Q - \delta x_P \\ \delta y_Q - \delta y_P \\ \delta z_Q - \delta z_P \end{bmatrix} + K \begin{bmatrix} \delta\Phi \\ \cos\phi\delta\Lambda \end{bmatrix} \tag{3-42}$$

where

$$K = \begin{bmatrix} \sigma\cos\alpha & \sigma\sin\alpha \\ \sigma\sin\alpha\cos\zeta & \sigma(\tan\phi\sin\zeta - \cos\alpha\cos\zeta) \\ 0 & 0 \end{bmatrix} \ ; \tag{3-43}$$

the matrices Q and S are given by (3-24) and (3-34).

These examples will illustrate how the linearized equations (3-8) can be obtained. Similar linearizations can be found in (Eeg and Krarup, 1975) and (Grafarend, 1977).

<u>Satellite observations</u>. Satellite observations can be linearized in the same way, as outlined in (Kaula, 1967,p.67) or (Heiskanen and Moritz, 1967, pp.352 - 355). This is theoretically

straightforward but computationally very laborious. For direction, distance, and Doppler measurements, linearizations can be found in many forms in the literature. Therefore, we need not go into details here. We only mention, that the linear functional LT is, for satellite observations, usually expressed in spherical harmonics:

$$LT = \sum_{n=2}^{\infty} \sum_{m=0}^{n} (a_{nm}\delta C_{nm} + b_{nm}\delta S_{nm}) , \tag{3-44}$$

which is a linear combination of the spherical-harmonic coefficients δC_{nm} and δS_{nm} of T.

For second-order gradients the linearization is straightforward: we have

$$\frac{\partial^2 V}{\partial x_i \partial x_j} = \left(\frac{\partial^2 V_E}{\partial x_i \partial x_j}\right)_0 + \left(\frac{\partial^3 V_E}{\partial x_i \partial x_j \partial x_k}\right)_0 \delta x_k + \frac{\partial^2 T}{\partial x_i \partial x_j} , \tag{3-45}$$

which is of form (3-5), V_E being the normal (ellipsoidal) gravitational potential. The vector δx_k can be further expressed in terms of the orbital parameters; cf. eq.(9-35) of (Heiskanen and Moritz, 1967, p.352).

A linearization for satellite-to-satellite tracking can be found in (Kryński, 1978).

4. Determination of the Gravity Field as an Improperly Posed Problem

A problem is called properly posed if the solution satisfies the following three requirements:

(1) existence,
(2) uniqueness,
(3) stability.

This means that a solution must exist for arbitrary (within a certain range) data, that there must be only one solution, and that this solution must depend continuously on the data. If one or more of these requirements are violated, then we have an improperly posed, or ill-posed, problem.

For a long time it was thought that only properly posed problems are physically meaningful. In fact, deterministic processes, as considered in classical mechanics, depend uniquely and continuously on the initial data--this is the essence of causality--and thus correspond to properly posed problems.

Only relatively recently is was recognized that there are important problems that are not properly posed. There is now an extensive literature on improperly posed problems; we mention only two easily accessible books: (Lavrentiev, 1967) and, especially, (Tikhonov and Arsenin, 1977), and the review article (Nashed, 1974). Geodetic applications are considered in (Schwarz, 1978); for instance, the downward continuation of gravity is an ill-posed problem. Also the work by Neyman (1977) should be mentioned.

Our present task, the determination of the earth's gravitational field from measurements, is a typical improperly posed problem. The potential is so irregular that it cannot be completely described by any finite set of parameters; on the other hand, we have only a finite number of measurements. Hence, there is no unique solution, and Condition 2 is violated.

Using the standard notation of (Tikhonov and Arsenin, 1977), we may write equations (3-8) in the form

$$Az = u \,, \tag{4-1}$$

if we put

$$u = \begin{bmatrix} \delta l_1 \\ \delta l_2 \\ \vdots \\ \delta l_q \end{bmatrix} \tag{4-2}$$

and

$$z = \begin{bmatrix} \delta X \\ T \end{bmatrix} . \tag{4-3}$$

A is a linear operator, expressing the fact that equations (3-8) permit the computation of the observations δl_k if δX and T are supposed to be given; the operations involved in these computations are clearly linear. The operator A, therefore, comprises the vectors a_k^T and the linear functionals L_k.

In mathematical terms, δX is a vector belonging to m-dimensional Euclidean space R_m, and T can be considered a member of a Hilbert space H of harmonic functions; u belongs to R_q. Therefore, the linear operator A (4-1) defines a linear mapping:

$$A: R_m \times H \rightarrow R_q . \tag{4-4}$$

The solution, if it exists, may be written formally in the form

$$z = A^{-1} u , \tag{4-5}$$

but it will certainly not be unique. Therefore, A^{-1} is not an inverse operator in the usual sense; it has the character of a generalized inverse operator (analogous to generalized matrix inverses).

For the solution of our problem we shall try to apply standard mathematical techniques for ill-posed problems. Nashed (1974, p.295) mentions the following possible approaches:

(a) a change of the concept of a solution;
(b) a change of the spaces and/or topologies;
(c) a change of the operator itself;
(d) the concept of regularization operators;
(e) probabilistic methods or well-posed stochastic extensions of ill-posed problems.

These approaches may overlap in various ways.

We shall especially use the approaches (b), (d), and (e); they will be considered in the following sections 5,6, and 7.

We finally point out that the vector X will only have linearly independent components and that also all equations of the system (3-8) are supposed to be linearly independent. In other terms, we shall work with systems (4-1) that have full rank.

5. Pure Collocation

Let us suppose that the systematic parameters X (coordinates, etc.) are known with sufficient accuracy. We then have $\delta X = 0$, and the system (3-8) reduces to

$$L_j T = l_j \ , \qquad j = 1,2,\ldots,q \ ; \tag{5-1}$$

for simplification, we have replaced δl_k by l_k. In order to find a solution, we apply Approach (b) mentioned at the end of the preceding section: a *change* of the solution space.

We shall try to approximate the desired function T by a linear combination f of suitable linearly independent base functions $\phi_1, \phi_2, \ldots, \phi_q$:

$$T(P) \doteq f(P) = \sum_{k=1}^{q} b_k \phi_k(P) \ , \tag{5-2}$$

P denoting the space point at which these functions are being considered, and b_k denoting suitable coefficients. Since T is harmonic outside the earth's surface, the base functions ϕ_k must also be harmonic functions.

Since there are q independent functions ϕ_k, the space of linear combinations (5-2) is q-dimensional, so that the solution space, as well as the observation space, is q-dimensional, and the operator A, consisting of the q functionals L_k, must, for this particular case, reduce to an $q \times q$ matrix, which in general has a regular inverse; hence the problem will, in general, become well-posed.

Note that this is made possible by <u>changing the solution space from infinitely-dimensional Hilbert space to the q-dimensional space of linear combinations</u> (5-2).

For the special case of interpolation, details can be found in (Moritz, 1978b, c).

We substitute (5-2) into (5-1) and get, because of linearity,

$$\sum_{k=1}^{q} b_k L_j \phi_k = l_j \quad . \tag{5-3}$$

The quantities

$$L_j \phi_k = A_{jk} \quad , \tag{5-4}$$

being the values of linear functionals, are constants. Hence we get the system of q linear equations

$$\sum_{k=1}^{q} A_{jk} b_k = l_j \quad , \tag{5-5}$$

which can be solved for the q coefficients b_k provided the determinant of the matrix A_{jk} is non-zero. The substitution of the b_k into (5-2) gives the desired solution.

The approximation of a function by fitting a linear combination of base functions to a number of linear functionals is <u>collocation</u> in the sense of approximation theory, cf.(Collatz, 1966, pp.29).

As a matter of fact, <u>interpolation</u> is a special case of collocation, when

$$L_j T = T(P_j) \quad , \tag{5-6}$$

that is, when L_j associates to a function its value at a particular point P_j; this is the so-called <u>evaluation functional</u>.

<u>Kernel functions</u>. Consider now a positive-definite symmetric function $K(P,Q)$, harmonic as a function of both P and Q. The positive definiteness of this function means that for any N, the $N \times N$ matrix of elements $K(P_j,P_k)$, for any points P_1, P_2, ..., P_N, is positive definite. Harmonicity means that Laplace's

equation is satisfied both at the point P (with Q held fixed) and at Q (for fixed P):

$$\Delta_P K(P,Q) = 0 \quad , \tag{5-7}$$

$$\Delta_Q K(P,Q) = 0 \quad ; \tag{5-8}$$

the domain of harmonicity may be the exterior of a certain sphere completely inside the earth.

Such a function $K(P,Q)$ will be called a <u>kernel function</u>.

Assume now that base functions ϕ_k have the special form

$$\phi_k(P) = L_k^Q K(P,Q) \quad , \tag{5-9}$$

where L_k^Q means that the linear functional is applied to the variable Q.

Then the matrix (5-4) takes the form

$$C_{jk} = L_j^P L_k^Q (K,Q) \quad , \tag{5-10}$$

and (5-5) becomes

$$\sum_{k=1}^{q} C_{jk} b_k = l_j \quad . \tag{5-11}$$

We solve this linear system and substitute the result into (5-2), which by (5-9) has the form

$$f(P) = \sum_{j=1}^{q} b_j \phi_j(P) = \sum_{j=1}^{q} b_j L_j^Q K(P,Q) \quad . \tag{5-12}$$

The result may be written as

$$f(P) = \begin{bmatrix} C_{P1} & C_{P2} & \cdots & C_{Pq} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1q} \\ C_{21} & C_{22} & \cdots & C_{2q} \\ \vdots & & & \vdots \\ C_{q1} & C_{q2} & \cdots & C_{qq} \end{bmatrix}^{-1} \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_q \end{bmatrix} \quad , \tag{5-13}$$

using the abbreviation

$$C_{Pj} = L_j^Q K(P,Q) \ . \qquad (5\text{-}14)$$

This is the basic formula for <u>analytical collocation with kernel functions</u>.

Interpolation with kernel functions is the special case (5-6), so that simply

$$\phi_j(P) = K(P,P_j) = C_{Pj} \ , \qquad (5\text{-}15)$$

$$C_{jk} = K(P_j,P_k) \ . \qquad (5\text{-}16)$$

This interpolation method is known from mathematics (Meschkowski, 1962, p.114); collocation with kernel functions has been applied in geodesy first by Krarup (1968, 1969).

A remark on terminology: the approximating function is obtained by fitting the given functionals exactly (measuring errors are not taken into account); therefore one speaks of "exact" or "pure" collocation. Sometimes it is also called "deterministic", to distinguish it from stochastic methods (Dermanis, 1976, p.56; Moritz, 1976, p.25), but the name "exact", or "pure", or "analytical", collocation seems to be better because the concept "determinism" already has a well-established meaning, namely causality in the sense of classical physics.

<u>Minimum norm property</u>. The solution (5-13) has an important property: among all possible solutions (in some Hilbert space of harmonic functions) of the system (5-1), the solution (5-13) has the minimum norm:

$$\| f \| = \text{minimum} \ , \qquad (5\text{-}17)$$

if the norm is defined by the inner product

$$\| f \|^2 = (f,f) \qquad (5\text{-}18)$$

in the Hilbert space with the kernel function $K(P,Q)$, so that there holds the reproducing kernel property

$$(f(P),K(P,Q)) = f(Q) , \tag{5-19}$$

where $(\cdot,\cdot)$ denotes the inner product with respect to P (Meschkowski, 1962, p.115; Krarup, 1969, p.39; Tscherning, 1975; Moritz, 1978b, p.36).

6. Application of Tichonov Regularization

Since we shall use the letter A for a different purpose later in this section, we shall denote the operator in (4-1) by G and write

$$Gz = u \quad . \tag{6-1}$$

Tichonov's regularization method consists in minimizing the nonlinear functional

$$M^{\alpha}[z,u] = ||Gz - u||^2 + \alpha\Omega(z) \quad , \tag{6-2}$$

where α is a numerical parameter and $\Omega(z)$ is a so-called <u>stabilizing functional</u> (Tichonov and Arsenin, 1977, pp.51,57), which may be taken as the square of some norm.

$$\Omega(z) = ||z||^2 \tag{6-3}$$

(*ibid*., p.72). In this way, a unique solution can usually be obtained.

Using (6-3) and very slightly generalizing (6-2) by the introduction of a second numerical parameter β, we get the condition

$$\alpha \|z\|^2 + \beta \|Gz - u\|^2 = \text{minimum} \quad . \tag{6-4}$$

Obviously, $Gz - u$ is the error in satisfying (6-1), so that $\|Gz - u\|$ will be the error norm. Therefore, (6-4) minimizes a linear combination of the norm of the solution z and of the error norm. Depending on the relation between α and β, a stronger weight is given to one or the other of these two norms.

The linear operator equation (6-1) is, in our case, nothing else than the system (3-8). Let us simplify the notation by replacing

$$\delta l_k \text{ by } l_k \ , \quad \delta X \text{ by } X \ . \tag{6-5}$$

Then (3-8) becomes

$$\begin{aligned} l_1 &= a_1^T X + L_1 T \ , \\ l_2 &= a_2^T X + L_2 T \ , \\ &\vdots \\ l_q &= a_q^T X + L_q T \ . \end{aligned} \tag{6-6}$$

where X is a m -vector (an $m \times 1$ matrix). We finally put

$$l = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_q \end{bmatrix} , \quad A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_q^T \end{bmatrix} , \tag{6-7}$$

and

$$B = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_q \end{bmatrix} . \tag{6-8}$$

Here l is a q -vector (a $q \times 1$ matrix), A is a $q \times m$ matrix, and B is a linear operator, formed of the q linear functionals L_k .

With these notation, (6-6) becomes

$$l = AX + BT \quad , \tag{6-9}$$

which is (6-1) with

$$u = l \quad , \tag{6-10}$$

$$z = \begin{bmatrix} X \\ T \end{bmatrix} \quad , \tag{6-11}$$

$$G = \begin{bmatrix} A & B \end{bmatrix} \; ; \tag{6-12}$$

we have used for linear operators the same mode of partitioning as for matrices. In fact, (6-1) becomes on partitioning

$$\begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X \\ T \end{bmatrix} = l \quad , \tag{6-13}$$

which is (6-9).

What is $Gz - u$ in our present case? We have

$$\begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X \\ T \end{bmatrix} - l = AX + BT - l \quad , \tag{6-14}$$

which is the amount by which the exact equation (6-1) or (6-9) is not satisfied. Let us denote this "misclosure" by $-n$:

$$AX + BT - l = -n \quad . \tag{6-15}$$

On rearrangement this becomes

$$l = AX + BT + n \quad ; \tag{6-16}$$

here n may be interpreted as the effect of measuring errors; we call n the "noise". It should be repeated that this equation is nothing but the system (3-8), obtained by linearizing the nonlinear functional equations (3-1).

Then, in (6-4),

$$||Gz - u|| = ||-n|| = ||n|| \ , \tag{6-17}$$

which is the norm of the q -vector n. Any regular quadratic norm in q -dimensional vector space can be written

$$||n||^2 = n^T Qn \tag{6-18}$$

with a positive definite regular symmetrix $q \times q$ "weight matrix" Q . Let us denote its inverse by D, then (6-18) becomes

$$||n||^2 = n^T D^{-1} n \ . \tag{6-19}$$

If n are random quantities in a statistical sense, then D may be considered as the covariance matrix of the measuring errors n. This statistical interpretation of n is not necessary--we may consider the norm (6-18) as a metric in a purely geometric sense--but it gives a clue as to the proper choice of the metric for the vector X. If the variance of a random quantity is small, then this quantity can vary only within narrow limits. The larger the variance, the larger variations are possible; and if the variance goes to infinity, the variation of our quantity becomes completely free.

Since we allow our parameter X to vary freely and independently, each component should have an infinite variance, or zero weight. Therefore, in an expression such as (6-18), with X instead of n, there should be $Q = 0$, which gives

$$||X|| = 0 \ . \tag{6-20}$$

This will be our choice for the norm of X.

Finally, the norm for T will be selected a norm in a Hilbert space with kernel function $K(P,Q)$, that is, defined by (5-18) and (5-19):

$$\|T\|^2 = (T,T) \ . \tag{6-21}$$

Then, assuming X and T independent of each other:

$$\|z\|^2 = \|X\|^2 + \|T\|^2 = \|T\|^2 \ . \tag{6-22}$$

Therefore, the Tichonov condition (6-4) becomes

$$\alpha\|T\|^2 + \beta\|n\|^2 = \textit{minimum} \tag{6-23}$$

or

$$\alpha(T,T) + \beta n^T D^{-1} n = \text{minimum} \ . \tag{6-24}$$

Pure collocation. As a preparation, let us assume errorless observations. Then $n = 0$ and (6-24) reduces to

$$(T,T) = \text{minimum} \tag{6-25}$$

(we have put $\alpha = 1$ without loss of generality). We furthermore assume $X = 0$ (no systematic effects). Then (6-16) gives

$$BT = l \tag{6-26}$$

where l is given. The desired T is that function T, satisfying (6-26), which minimizes (6-25).

We solve the problem by means of a Lagrange multiplyer. Instead of minimizing (6-25) under the side condition (6-26), we form the unconditional minimum of the function

$$\Phi = \frac{1}{2}(T,T) - k^T(BT - l) \ , \tag{6-27}$$

where the q -vector k serves as a Lagrange multiplyer.

A necessary condition for a minimum is the vanishing of the differential of Φ:

$$d\Phi = (T,dT) - k^T B dT = 0 \quad . \tag{6-28}$$

(We can form this differential as if T were a vector. To avoid misunderstanding, we point out, however, that dT is not an ordinary differential of T, but what is called, in the calculus of variations, a first variation, that is, a change in the function T. In fact, (6-24) is a variational principle, and (6-28) is the corresponding Euler equation for our special case. Eq. (6-28) may be found by replacing, in (6-27), T by $T + \varepsilon\tau$, where ε is a small parameter:

$$\Phi_\varepsilon = \frac{1}{2}(T + \varepsilon\tau, T + \varepsilon\tau) - k^T(BT + \varepsilon B\tau - 1) =$$

$$= \frac{1}{2}(T,T) - k^T(BT - 1) + \frac{1}{2}\varepsilon(T,\tau) + \frac{1}{2}\varepsilon(\tau,T) -$$

$$- \varepsilon k^T B\tau + \frac{1}{2}\varepsilon^2(\tau,\tau) \quad .$$

By symmetry, $(\tau,T) = (T,\tau)$. We subtract (6-27) and divide by ε. On letting $\varepsilon \to 0$, we thus get

$$\lim_{\varepsilon\to 0}\frac{\Phi_\varepsilon - \Phi}{\varepsilon} = (T,\tau) - k^T B\tau = 0 \ , \tag{6-29}$$

which is (6-28), with $dT = \varepsilon\tau$.)

The function dT in (6-28) is completely arbitrary; it need not even be small since a numerical factor does not matter. (Of course, dT must belong to the Hilbert space under consideration.)

By the reproducing property (5-19),

$$dT(Q) = (dT(P), K(P,Q)) \tag{6-30}$$

or, briefly,

$$dT = (dT,K) = (K,dT) \ . \tag{6-31}$$

Hence,

$$BdT = (BK,dT) \quad , \tag{6-32}$$

and (6-28) becomes

$$(T,dT) - (k^T BK,dT) = 0$$

or

$$(T - k^T BK,dT) = 0 \ . \tag{6-33}$$

Since dT is arbitrary, there must be

$$T - k^T BK = 0$$

or

$$T = k^T BK \ . \tag{6-34}$$

This is an important result. What does it mean? In view of (6-8) and (6-30), this is nothing else than

$$T(Q) = \sum_{j=1}^{q} k_j L_j^P K(P,Q) \ ; \tag{6-35}$$

L_r^P means the operator L_r applied to the variable P. Now, (6-35) is identical to (5-12), with P and Q interchanged and $b = k_j$. Thus, <u>the best approximation for</u> $T(Q)$ <u>is</u>, in fact, <u>a linear combination of the base functions (5-9)</u>!

The rest is straightforward. Considering T a scalar, we may transpose (6-34):

$$T = (BK)^T k \quad , \tag{6-36}$$

and substitute into (6-26):

$$B(BK)^T k = 1 \ . \tag{6-37}$$

The $q \times q$ matrix

$$C = B(BK)^T \tag{6-38}$$

has, by (6-8), the elements

$$C_{ij} = L_i^P L_j^Q K(P,Q) \ . \tag{6-39}$$

Now (6-37) may be solved for k:

$$k = C^{-1}l \ , \tag{6-40}$$

so that (6-36) becomes

$$T = (BK)^T C^{-1} l \ . \tag{6-41}$$

In view of (5-14), this is identical to (5-13).

We thus have obtained (5-13) as a consequence of the minimum norm principle (6-25). There are shorter and more complete proofs (the condition (6-28) is necessary but not sufficient); the advantage of the present derivation is the treatment as a straightforward solution of a variational principle by standard techniques (Euler equation). Furthermore, it will essentially simplify the treatment of the general case.

<u>The case</u> $\alpha = \beta = 1$. As a second step, let us consider the general equation (6-16), but put, in the Tichonov condition (6-24), $\alpha = \beta = 1$, so that

$$(T,T) + n^T D^{-1} n = \text{minimum} \ , \tag{6-42}$$

to be solved under the side condition (6-16). We thus have to form the unconditional minimum of the function

$$\Phi = \frac{1}{2}(T,T) + \frac{1}{2}n^T D^{-1} n - k^T(AX + BT + n - l) \ . \tag{6-43}$$

The differential is

$$d\Phi = (T,dT) + n^TD^{-1}dn - k^T(AdX + BdT + dn) \ , \qquad (6\text{-}44)$$

where dX and dn are ordinary vector differentials. On rearranging we get, using (6-32),

$$d\Phi = (T - k^TBK,dT) + (n^TD^{-1} - k^T)dn - k^TAdX = 0 \ . \qquad (6\text{-}45)$$

Since dT, dn, and dX are arbitrary, $d\Phi = 0$ can only hold if

$$T - k^TBK = 0 \ , \qquad (6\text{-}46)$$

$$n^TD^{-1} - k^T = 0 \ , \qquad (6\text{-}47)$$

$$k^TA = 0 \ . \qquad (6\text{-}48)$$

The first equation gives

$$T = k^TBK \ , \qquad (6\text{-}49)$$

identical to (6-34) or (6-35). <u>Again, the solution is a linear combination of base functions</u> (5-9)!

The transposition of (6-49) gives

$$T = (BK)^Tk \ , \qquad (6\text{-}50)$$

so that

$$BT = B(BK)^Tk = Ck \ , \qquad (6\text{-}51)$$

using the abbreviation (6-38). Eq.(6-47) gives

$$n^T = k^TD \quad \text{or} \quad n = Dk \ . \qquad (6\text{-}52)$$

Eq.(6-16) may be written

$$1 - AX = BT + n \,, \tag{6-53}$$

and substituting (6-51) and (6-52) we get

$$1 - AX = (C + D)k \,, \tag{6-54}$$

so that

$$k = (C + D)^{-1}(1 - AX). \tag{6-55}$$

We substitute this into (6-48), transposed as

$$A^T k = 0 \,,$$

obtaining

$$A^T(C + D)^{-1}1 - A^T(C + D)^{-1}AX = 0 \quad ,$$

so that

$$X = \left[A^T(C + D)^{-1}A\right]^{-1} A^T(C + D)^{-1}1 \,, \tag{6-56}$$

which determines the parameter vector X. The substitution of (6-55) into (6-50) then gives the potential:

$$T = (BK)^T(C + D)^{-1}(1 - AX) \,. \tag{6-57}$$

<u>The general case</u>. Take finally the general Tichonov condition (6-24)

$$\alpha(T,T) + \beta n^T D^{-1} n = \text{minimum} \tag{6-58}$$

for solving the equation (6-16)

$$AX + BT + n = 1 \,. \tag{6-59}$$

The condition (6-58) is, for $\alpha \neq 0$, equivalent to

$$(T,T) + n^T(\frac{\alpha}{\beta}D)^{-1}n = \text{minimum} \ , \tag{6-60}$$

so that we only have to replace, in (6-56) and (6-57), D by $\alpha D/\beta$. This gives

$$X = \left[A^T(\beta C + \alpha D)^{-1}A\right]^{-1} A^T(\beta C + \alpha D)^{-1}l \ , \tag{6-61}$$

$$T = (\beta BK)^T(\beta C + \alpha D)^{-1}(l - AX) \ , \tag{6-62}$$

which is the solution of (6-59) under the general Tichonov condition (6-58).

Obviously, to various ratios $\alpha : \beta$ there corresponds a different weighting between the square of the "function norm", (T,T), and of the "error norm", $n^TD^{-1}n$. Pure collocation, with the condition (6-25) and $n = 0$, fits the solution exactly to the data. This is unsuited for real data in the presence of measuring errors, because then the solution is distorted by faithfully reproducing all measuring errors; we risk to get spurious oscillations (Eeg and Krarup, 1975, p.111).

Therefore, some balance between α and β in (6-58) must be found. But how? We might use some trial-and-error procedure, but this does not seem very satisfactory. A theoretically motivated, in a certain sense optimal, solution to this problem is found by statistical considerations, as we shall see in the following section.

In their "Integral Geodesy", Krarup and Eeg (1975) give equations equivalent to (6-61) and (6-62), but with different weighting in the two equations: in (6-61) they use $\beta = 1 - \alpha$ (ibid., p.119, their eq.(5)) and in (6-62) they use $\beta = \alpha = 1$ (ibid., p.112, their eq.(14)). This follows from employing two different minimum principles. It seems, however, preferable to derive both equations from the same variational principle (6-58).

We also mention that the result for the "errorless" condition (6-25) cannot be obtained by simply putting $\alpha = 1$, $\beta = 0$, as might be expected at first sight. The essential feature with (6-25) is that $n = 0$ and $D = 0$. We, therefore, have to put $D = 0$ in (6-61) and (6-62). As a consequence, β cancels then, and we obtain

$$X = (A^T C^{-1} A)^{-1} A^T C^{-1} l \ , \tag{6-63}$$

$$T = (BK)^T C^{-1} (l - AK) \ . \tag{6-64}$$

For $A = 0$ (no systematic parameters), the last equation reduces to (6-41), as it should.

A final word on the solution of these variational principles by an Euler equation. Any Euler equation gives only a necessary, not a sufficient, condition for a minimum. It is not too difficult to show that our solutions do indeed give a minimum. The proof goes along well-known lines (cf. Moritz, 1972, pp.22-23).

7. Least-Squares Collocation

One of the possible approaches to improperly posed problems mentioned at the end of sec.4 are stochastic methods. We have already been close to such an approach in the preceding section, when we interpreted the matrix D in (6-19) as the covariance matrix of the measuring errors; however, this was not essential since, from an analytical point of view, we could as well have used any other positive-definite regular square matrix.

A more thoroughgoing stochastic approach consists in the attempt to interpret the norm $\|T\|$ statistically, as well as the error norm $\|n\|$. This can be achieved by formally considering the anomalous gravity potential T as a stochastic process. This could be a purely formal "stochastic extension", motivated by the irregularity of the anomalous gravity field.

Historically, this was the way that led to collocation in its geodetic application. The statistical treatment of the gravity anomalies was initiated by DeGraff-Hunter in 1935, but did not find much response. The contemporary rapid development started from Hirvonen's work, published in 1956: the explicit treatment of the anomalous gravity field as a stochastic process in 1959 by Kaula, the geodetic application of least-squares prediction techniques of stochastic processes by Moritz in 1962, generalized by Kaula in 1963, and the extension to the estimation of T from given functionals, which is collocation in the proper sense, by Krarup in 1968-9.

Stochastic process techniques form a very convenient tool and an intuitive terminology (variances, covariance functions, etc.), and they permit the estimation of accuracies, which is of particular importance for geodetic applications.

There is no *fundamental* objection against embedding the actual Earth, with its potential T, in an ensemble of "sample earths", making possible the application of stochastic-process techniques. Similar approaches are frequently applied (consciously or unconsciously) in various other fields.

From the point of view of logical simplicity, however, it appears preferable to find an interpretation, which retains the convenient formal apparatus of stochastic processes, but is restricted to our Earth only, without introducing fictitious other "earths".

This is made possible by an application of Norbert Wiener's "covariance analysis of individual functions", in our case, of the actual anomalous gravity potential T (Moritz, 1972, sec.8).

In this case, the covariance function of T is defined as

$$C(P,Q) = M_1\{T_P T_Q\} , \tag{7-1}$$

as a homogeneous and isotropic average M_1 of the product $T_P T_Q$, where

$$M_1\{\cdot\} = \frac{1}{8\pi^2}\int_{\lambda=0}^{2\pi}\int_{\theta=0}^{\pi}\int_{\alpha=0}^{2\pi}(\cdot)\sin\theta d\theta d\lambda d\alpha \tag{7-2}$$

is an average over all points (θ,λ) and all azimuths α (ibid., p.114), which can be considered as an average over rotation group space.

If we consider rotation group space as a probability space and on this space a stochastic process, the sample functions of which differ only by rotations, then the probability average coincides, by definition, with the "space average" (7-1), so that our process will be (trivially) ergodic (Moritz, 1972, p.119). This is elaborated in detail in (Moritz, 1978a), where ergodic processes on the sphere are discussed, They are of necessity non-Gaussian, since Lauritzen has shown in 1971 that ergodic Gaussian processes on the sphere are impossible.

The statistical expectation in the probability space of the measuring errors n are denoted by M_2.

Thus the measuring errors may be considered as "physically stochastic" quantities. The anomalous potential T is only formally a space average. If we understand statistics as the study of a large amount of data and their average properties (for instance, an analysis of the global human population), then we may say that our treatment of T is statistical, though T itself is not a stochastic quantity in a strictly physical sense (Moritz and Sansò, 1978).

Least-squares collocation is defined as that collocation method which minimizes the variance of the estimated quantities, in our case, of the potential T and the parameter vector X. Let T^* be the true potential and T its estimate from the given data, affected by measuring errors, at a particular point P. Then the variance of the estimate T, or the square of the prediction error m_P, is defined as

$$m_P^2 = M_1 M_2\{(T_P^* - T_P)^2\} , \tag{7-3}$$

that is, as the average of the square of the true prediction error $T_P^* - T_P$ at some point P, extended over all measuring errors (M_2) and all points of the sphere (M_1).

This definition is very natural; it corresponds to a global average.

Similarly the variance m_i^2 of the estimate for a component X_i of the parameter vector X is defined. The conditions

$$m_P^2 = \text{minimum} , \qquad (7\text{-}4)$$

$$m_i^2 = \text{minimum} , \qquad (7\text{-}5)$$

for all points P of the sphere and all components of X, define least-squares collocation.

Special cases are least-squares prediction of gravity, which uses only (7-4)--cf. (Heiskanen and Moritz, 1967,p.268)-- and, of course, least-squares adjustment, which uses only (7-5).

The result of conditions (7-4) and (7-5) are as follows: we get collocation in a Hilbert space, in which the kernel function is given by the covariance function (7-1), and T and X are then expressed by (6-56) and (6-57). Putting

$$C + D = C_{xx} = \overline{C} , \qquad (7\text{-}6)$$

$$BK = C_{sx} , \qquad (7\text{-}7)$$

we get the well-known formulas of least-squares collocation

$$X = (A^T\overline{C}^{-1}A)^{-1}A^T\overline{C}^{-1}l , \qquad (7\text{-}8)$$

$$s = C_{sx}\overline{C}^{-1}(l - AX) . \qquad (7\text{-}9)$$

Here we have replaced T by s, because the anomalous potential, or any other quantity of the anomalous gravitational field, is also called a "signal", denoted by s. A pleasant consequence of least-squares methods with respect to linear transformation is that an equation of form (7-9) holds for the direct estimation

of any anomalous field quantity, all these quantities (signals) being linear functions of T. In fact, eq.(2-36) of (Moritz, 1972), has been derived for the estimation of any signal.

Thus, least-squares collocation corresponds to a Tichonov condition (6-58) for which the kernel function defining the norm $\sqrt{(T,T)}$ is the covariance function, and for which $\alpha = \beta = 1$. Thus, the relative weights of the two summands (6-58) are uniquely determined in such a way as to obtain an optimal (minimum variance) result.

The observation equation to be satisfied is (6-16):

$$l = AX + BT + n \; . \tag{7-10}$$

Putting

$$BT = s' \; , \tag{7-11}$$

which is the influence of the anomalous field on l, that is, the "signal part" of l, we have

$$l = AX + s' + n \; , \tag{7-12}$$

which is the fundamental equation of least-squares collocation; cf. eq.(2-1) of (Moritz, 1972, p.7), with l denoted by x.

The simplest way to derive (7-8) and (7-9) is to use a finite minimum norm principle:

$$s^T C^{-1} s + n^T D^{-1} n = \text{minimum} \; , \tag{7-13}$$

where the vector s comprises the components of s' in (7-12) plus the signals to be estimated in the given problem (and only these); C is the covariance matrix of this vector s; cf. (Moritz, 1972, p.122); a similar principle is used (ibid., sec.2).

This approach is completely equivalent to (6-42), which uses the full Hilbert space norm. Here only finite matrix operations are required; Hilbert space lurks invisibly in the background (to use another metaphor: matrix formulas such as (7-8) and (7-9) represent the finite-dimensional surface of infinitely-dimensional Hilbert space).

Thus, least-squares collocation satisfies two conditions: a minimum norm condition ((6-42) or (7-13)) and a minimum variance condition ((7-4) and (7-5)).

The minimum norm condition (especially in the form (7-13)) is mathematically convenient to handle, but physically it appears somewhat unnatural: it combines additively two different quantities (norms), which have a completely different physical character: the first depends on the anomalous gravitational field, the second on random measuring errors.

Much more natural seems the minimum variance condition: the average in (7-3) (and similarly in (7-5)) is over both signal and noise, but affecting the same quantity $(T_P^* - T_P)^2$: this is felt to be exactly as it should be.

An interesting variant of the minimum variance approach is to start from a principle such as (7-4) and to postulate a linear estimation that is symmetric with respect to the rotation group. In this way it is possible to deduce (7-1), rather than introducing it by definition (Sansò, 1978).

Error Covariances. Probably the most important feature of the statistical approach is the possibility to estimate, not only the values of anomalous field quantities and of systematic parameters, but also their accuracy.

The error variances of the parameter vector X and the signal s and their covariances are expressed by the error covariance matrices E_{XX}, E_{ss} and E_{Xs}, the last being the cross-covariance matrix of X and s. The vector s comprises any number of signals (anomalous field quantities) estimated by the least-squares collocation procedure under consideration.

We have (Moritz, 1972, pp.30-33)

$$E_{xx} = (A^T \overline{C}^{-1} A)^{-1} , \quad (7\text{-}14)$$
$$E_{ss} = C_{ss} - C_{sx} \overline{C}^{-1} C_{xs} + HAE_{xx} A^T H^T , \quad (7\text{-}15)$$
$$E_{xs} = -E_{xx} A^T H^T , \quad (7\text{-}16)$$

where

$$H = C_{sx} \overline{C}^{-1} . \quad (7\text{-}17)$$

The matrix C_{ss} is the signal covariance matrix of the s that was to be estimated; E_{ss} is its error covariance matrix.

These formulas express the accuracy of the statistically optimal estimates (7-8) and (7-9). The statistical treatment of the gravity field, however, makes it even possible to estimate the accuracy of other linear estimation formulas such as (6-61) and (6-62). We write these equations in the form

$$X = \hat{G} l , \quad (7\text{-}18)$$
$$\begin{aligned} s &= \hat{H}(l - AX) = \hat{H}(l - A\hat{G}l) \\ &= \hat{H}(I - A\hat{G})l \\ &= \hat{L} l , \end{aligned} \quad (7\text{-}19)$$

I denoting the unit matrix and the vector s comprising those values of T that were estimated.

The least-squares estimates (7-8) and (7-9) may be written in an analogous form:

$$X = Gl , \quad (7\text{-}20)$$
$$s = Ll . \quad (7\text{-}21)$$

Then equations (3-52a,b) of (Moritz, 1972) give:

$$\hat{E}_{xx} = E_{xx} + (\hat{G} - G)\overline{C}(\hat{G} - G)^T , \quad (7\text{-}22)$$
$$\hat{E}_{ss} = E_{ss} + (\hat{L} - L)\overline{C}(\hat{L} - L)^T . \quad (7\text{-}23)$$

Here $\hat{E}_{xx}$ and $\hat{E}_{ss}$ are the error covariance matrices of the estimates (6-61) and (6-62), and E_{xx} and E_{ss} are given by (7-14) and (7-15). It should be noted that, in the least-squares estimates (7-8) and (7-9), the matrices C and C_{sx} are derived from the covariance function, whereas the corresponding matrices C and BK in (6-61) and (6-62) are to be derived from the kernel function used.

Geometrical interpretation of minimum variance. The condition of minimum variance has an interesting geometrical interpretation by means of the dual space, which is the space of all bounded linear functionals of the elements of the given linear space (Tscherning, 1978, p.175).

Let, for simplicity, be $A = 0$ and $D = 0$, that is, consider the case of pure collocation without random errors and systematic effects. Then (6-6) reduces to

$$l_i = L_i T , \tag{7-24}$$

and the least-squares estimate (7-9) becomes

$$s = C_{sx} C^{-1} l . \tag{7-25}$$

Here

$$s = FT \tag{7-26}$$

is a linear functional, to be estimated, of the anomalous potential T; as a matter of fact, (7-25) could also have been obtained by applying the linear functional F on both sides of eq. (5-13).

The estimate (7-25), besides satisfying the requirement of minimum norm $\|T\|$, also satisfies the condition

$$\|F^* - F\|' = \text{minimum}, \tag{7-27}$$

where $\|\cdot\|'$ denotes the norm in the dual space and F^* stands for the "true" value of the functional, of which F is the estimate. The minimum (7-27) is taken with respect to all possible estimates of F as linear combinations of the q given functionals L_i.

In geometrical terms, the "best" estimate (7-25) is simply the orthogonal projection, of the functional to be estimated, onto the subspace of the dual space spanned by the q given functionals L_i.

For the case of pure collocation, this is shown in (Krarup, 1978, p.200). (The geometrical interpretation is perhaps best understood in the simple finite-dimensional case of ordinary least-squares adjustment, where linear functionals reduce to vectors (Moritz, 1966).)

This geometrical situation is valid for arbitrary kernel functions. A physical interpretation, however, seems to be possible only if we identify the kernel function with the covariance function; then $\|F^* - F\|'$ is simply the standard error m_F of estimating the functional F. In fact,

$$(F^* - F)T = F^*T - FT = s^* - s = \varepsilon_F \qquad (7\text{-}28)$$

is the "true error", the difference between true value s^* and predicted value s, and

$$\| F^* - F\|'^2 = M\{\varepsilon_F^2\} = m_F^2 \qquad (7\text{-}29)$$

is the average of ε_F^2, that is, the square of the standard prediction error m_F.

Remark on Terminology. The minimum variance conditions (7-4) and (7-5) correspond to similar conditions in least-squares adjustment and least-squares prediction. Therefore, the name, least-squares collocation, seems to be appropriate for the case in which the solution is given by (7-8) and (7-9).

In contrast to this name, collocation using an arbitrary kernel function, which is the case with the methods described in sections 5 and 6, might be called collocation with kernel functions. Pure collocation (sec.5) would then be kernel function collocation without noise (and without systematic parameters), and the case of "integrated geodesy"--formulas such as (6-61) and (6-62)--would then be kernel function collocation with noise, including the estimation of systematic parameters.

It is evident that also the results of collocation with a general kernel function satisfy a minimum principle of form (6-24) and, dually, (7-27), which might also be called (geometrical) least-squares principles; therefore Krarup (1978, p.197) uses the term "least-squares collocation" also in the case of a general kernel function.

The terminology suggested here has the advantage of distinguishing between "statistical" and "geometrical" collocation.

We finally mention that a comprehensive review of quadratic norm and minimum variance estimation principles is given in (Grafarend, 1978). Grafarend also considers rank-deficients equations, occuring when the components of the vector X are not linearly independent.

8. Review; Alternatives

In this report we have asked ourselves in which way arbitrary geodetic observations of the earth's gravitational field can be used to obtain information about this field and to determine this field to the best possible extent. We have seen that, using contemporary mathematical techniques, a straight road leads from the nonlinear observational equations to collocation with kernel functions and least-squares collocation. The main stages on the way were linearization (sec.3) to obtain a linear improperly posed problem (sec.4), to which three different standard methods of solution are applied: a restriction of the solution space, leading to "pure collocation" (sec.5), variational principles of Tichonov type, by which measuring errors can be taken into account, leading to a generalized collocation with kernel functions, and a statistical approach, leading to least-squares collocation.

We shall now discuss these various stages and possible alternatives.

Linearization. All geodetic observations are nonlinear functionals of the potential V and of certain parameters X. Each observation gives a functional equation, and q observations give a set of q of such observations. The determination of V and X from this system of equations is a nonlinear improperly posed problem. By a Taylor linearization we get a system of q linear functional equations for the anomalous potential T and for corrections, again denoted by X, to the parameter vector. For this linear improperly posed problem, standard mathematical techniques exist. If the accuracy of linearization is not sufficient, we may iterate. A direct attack of the nonlinear problem does not seem possible with present mathematical tools: there is at present no alternative to linearization.

Why collocation? Collocation in a mathematical sense is the approximation of a function by fitting an analytical expression to q given linear functionals; for a very simple example see (Moritz and Sünkel, 1978, p.32).

Such *collocation methods* are used in applied mathematics for the approximate solution of differential equations, etc. (which are also functional equations!). Here the functionals are usually supposed to be given in a mathematically exact way, and the analytical expression is required to fit these data exactly.

Such a "pure" or "mathematical" collocation is, in general, not adequately applicable to our present geodetic problem, in view of the inevitable random measuring errors (noise). An exception is, for instance, least-squares interpolation of gravity anomalies (interpolation is a special case of collocation, in which the functionals are simply the values of the function at discrete points); here, the measuring errors of gravity are considered to be negligibly small. Generally, measuring errors, or "noise", must be suitably taken into account: we have a problem of "collocation with noise". (This is true for the linearized problem, but may be said to hold even for the original nonlinear problem because measurements, by their very nature, are nonlinear functionals of V and X, and the notion of collocation may be extended, in a natural way, also to the fitting of nonlinear functionals. But, as we have said, we shall anyway restrict ourselves here to linearized problems.)

Thus the operational approach to physical geodesy, starting from the measurements, inevitably lead to a collocation problem (with noise); there is, in this sense, no alternative to collocation. What can be chosen in different ways, is the analytical expression for approximating the potential.

Linear or nonlinear approximation. Instead of a linear combination (5-2) of base functions one could, in principle, also envisage other forms of approximation. For reasons of mathematical simplicity, and also in view of the smallness of the quantities under consideration (anomalous field quantities and corrections to the parameters), linear approximations are used practically without exception. (An exception would be nonlinear prediction to be mentioned later which is, however, hardly used in geodetic practice.) From the point of view of theoretical simplicity and practical usefulness, the restriction to linear approximation methods seems to be fully justified.

Alternative base functions. Besides kernel functions, many other functions are being used as base functions in expressions such as (5-2). We only mention polynomials in polynomial interpolation, trigonometric functions in trigonometric interpolation and approximation and, in geodetic applications, multiquadric functions (Hardy, 1976). Particularly useful are spline functions and other finite elements; for an elementary discussion and comparison with kernel functions cf. (Moritz, 1978b). An interesting geodetic application of spline functions is to a fast computation of covariance functions (Sünkel, 1978).

These functions can be very well suited for particular applications, but they cannot be used to solve the general operational problem of physical geodesy because they are not harmonic outside the earth, which is required for approximating the potential T.

Spherical harmonic functions can be used to represent the potential, and they are, in fact, fundamental for this purpose. From a practical point of view, they are particularly suited for representing the global field at satellite altitudes. The sample functions of Giacaglia and Lundquist (1972) are finite linear combinations of spherical harmonics in a form convenient for certain purposes. For local representations of the detailed gravity field, however, spherical harmonics are not applicable. This excludes their use as base functions in the present general context.

Kernel functions can equally well be used for local and global purposes; this explains their application in our general operation approach.

It should be mentioned also that any choice of base functions leads to a $q \times q$ system of linear equations for the coefficients; for kernel functions, this system will have a symmetric matrix.

Alternative norms. Let us now turn to the Tichonov principle (6-4). Why should we use a quadratic norm, of form (6-19) and (6-21), respectively? The reason is simple: only then will we get a linear variational (Euler) equation, leading to a linear combination of base functions.

This motivates the use of a quadratic norm, which is the inner product of T with itself:

$$\|T\|^2 = (T,T) \quad , \tag{8-1}$$

that is, the use of Hilbert space. But why a Hilbert space with kernel functions, not some other Hilbert space?

The reason is that only then the Euler equation will, in fact, lead to a linear combination (5-2) of base functions. Let us illustrate this by means of a counterexample. If we restrict ourselves to the interpolation of functions f defined on a sphere σ, then a very obvious choice of norm would be

$$\|f\|^2 = (f,f) = \iint_\sigma f^2 d\sigma \ . \tag{8-2}$$

The Hilbert space so defined does not have a kernel function in the proper sense; however, it may be considered as a Hilbert space with a "generalized" kernel function, which is a Dirac delta function defined by

$$\begin{aligned} &\delta(P,Q) = \delta(Q,P) \quad , \\ &\delta(P,Q) = 0 \quad \text{if} \quad P \neq Q \ , \\ &\iint_\sigma \delta(P,Q) d\sigma_P = 1 \quad \text{for fixed} \quad Q \ . \end{aligned} \tag{8-3}$$

Then

$$\begin{aligned} (f(P),K(P,Q)) &= (f(P),\delta(P,Q)) = \iint_\sigma f(P)\delta(P,Q)d\sigma_P = \\ &= f(Q)\iint_\sigma \delta(P,Q)d\sigma_P = f(Q) \ , \end{aligned}$$

in agreement with the definition (5-19).

For our interpolation on the sphere we thus have, by (5-15),

$$\phi_j(P) = \delta(P,P_j) \ . \tag{8-4}$$

These "functions" are zero outside the interpolation points. The interpolation function, the linear combination (5-2), has, therefore, the property that it is zero everywhere outside the points at which the functional values are given; there is no smooth interpolation.

This simple example will illustrate that only quadratic norms with kernel functions can be used.

It is not difficult to see that a kernel function norm is a natural extension, to Hilbert space, of the usual quadratic norm

$$x^T K^{-1} x \ , \tag{8-5}$$

with a positive definite regular matrix K, of vectors x in a finite-dimensional space.

Thus, the Tichonov approach with a quadratic norm inevitably leads to collocation with kernel functions; other base functionscannot be obtained in this way.

In order to further illustrate this, let us consider another example. Assume a function $y = f(x)$ to be defined on the unit circle $(0 \leq x \leq 2\pi)$, such that

$$f(2\pi) = f(0) \ . \tag{8-6}$$

Let us further assume that the function has a square-integrable second derivative $f''(x)$ and define the norm of f by

$$\| f \|^2 = \int_0^{2\pi} |f''(x)|^2 \, dx \tag{8-7}$$

(it is actually a semi-norm, but this is irrelevant here).

Expanding the function $f(x)$ into a trigonometric series we have

$$f(x) = \frac{1}{2} a_o + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) \quad , \tag{8-8}$$

$$f''(x) = - \sum_{n=1}^{\infty} n^2 (a_n \cos nx + b_n \sin nx) \quad . \tag{8-9}$$

The substitution of the last equation into (8-7) and subsequent integration leads to

$$|| f ||^2 = \pi \sum_{n=1}^{\infty} n^4 (a_n^2 + b_n^2) \quad . \tag{8-10}$$

Using as kernel function

$$K(P,Q) = \frac{1}{\pi} \sum_1^{\infty} n^{-4} \cos(x - \xi) \quad , \tag{8-11}$$

where

$$x_P = x \; , \qquad x_Q = \xi \; ,$$

we have

$$(f(P), K(P,Q)) = \int_0^{2\pi} \frac{d^2 f}{dx^2} \frac{\partial^2 K}{\partial x^2} dx \quad , \tag{8-12}$$

which is readily seen to be equal to $f(\xi)$, provided

$$a_o = \frac{1}{\pi} \int_0^{2\pi} f(x) dx = 0 \quad . \tag{8-13}$$

Thus, for functions satisfying (8-13), (8-11) is the reproducing kernel. Interpolation and collocation with such a kernel function correspond to minimizing the norm (8-7).

The minimum of a norm such as (8-7), the integral being extended over an interval $[a,b]$, is used to define cubic spline functions (there are nonperiodic and periodic splines); cf. (Moritz and Sünkel, 1978, pp.26 and 44).

The physical interpretation of minimum norm (8-7) is minimizing the total curvature. Similarly, we could, in geodesy, use the condition that the total mean curvature of the geoid (the integral, over the sphere, of the square of mean curvature) is minimized, in order to define a kernel function.

Least-squares collocation versus collocation with a general kernel function. In least-squares collocation, the kernel function is chosen to be the covariance function; see the end of sec.7. The basic advantage of this choice is the statistical significance: it gives the best linear estimate (minimum variance of the result).

Since the anomalous potential is not normally distributed (sec.7), the "best linear" estimate is not necessarily the absolutely "best" estimate: nonlinear estimates may still reduce the variance of the result. Therefore, nonlinear prediction has been considered (Kaula, 1966; Grafarend, 1972). For most practical purposes, however, linear estimates seem to be fully adequate.

Of decisive importance is the possibility, in least-squares collocation, to estimate the accuracy of the results. This can be done solely on the basis of the covariances, without needing actual measurements. Therefore, least-squares collocation can be used for investigating possible configurations of measurements, for the planning of surveys, and even for the investigation of accuracies to be expected with the use of measuring techniques which are being developed or only being envisaged for the future.

An interesting feature has been pointed out by Tscherning (1977): if we take the covariance function as kernel function, then the anomalous potential T has an infinite norm, so that T itself does not belong to the Hilbert space under consideration. In physical terms, taking for example interpolation, this means that the interpolating linear combinations of kernel functions are all much smoother than T itself. In practice, this is an advantage rather than a deficiency; it is, however, a slight theoretical drawback because it makes it difficult to prove convergence of the approximating to the "true" function if the density of the data approaches a continuous distribution (Tscherning, 1977). This drawback can easily be remedied by changing the higher-order spherical-harmonic terms of the covariance function (which can anyway not be exactly determined because of lack of data) by an arbitrarily small amount which does not noticeably change this function.

The exact covariance function could, of course, only be determined if we knew T exactly (and then we should not need collocation!). So all we can do is to fit an appropriate analytical expression to empirical data (gravity variances and covariances, degree variances from a spherical-harmonic development, etc.).

So, in the practice of least-squares collocation, we use an analytical kernel function which approximates the covariance function without coinciding with it. It should, however, be close enough so that accuracy estimates are meaningful.

Even the use of a general harmonic kernel function precisely preserves the mathematical structure of the gravity field, expressed by relations between the potential, gravity anomalies, deflections of the vertical, etc..

The approach followed in practice by most workers in this field, whether they explicitly favor a statistical interpretation or not, is to use a kernel function which has a simple analytical expression and exhibits the main features of an empirical covariance function.

Relation to least-squares adjustment. Least-squares collocation differs from least-squares adjustment in two respects.

First, from a physical point of view, adjustment contains only one kind of statistical quantity, namely the random measuring errors (noise); in least-squares collocation there are two physically different quantities that are treated statistically: quantities of the anomalous gravity field (the signal) and measuring noise.

Second, from a formal-mathematical point of view, the important distinction is that the signal, in contrast to the noise, is a continuous function; therefore, an adequate treatment requires some infinitely-dimensional function space, especially Hilbert space. Operations in Hilbert space show formal similarities with operations in finite-dimensional space but are qualitatively different, just as differential equations are qualitatively different from difference equations although there are formal similarities.

It is true that the gravity field at satellite elevations can be adequately described by a spherical-harmonic expansion truncated at a sufficiently high degree; the coefficients of such a development do form a finite-dimensional vector. Thus, if we exclusively work at satellite altitudes, we might, in fact, replace Hilbert space by a finite-dimensional space, without essentially impairing the accuracy.

This situation changes essentially if we consider the gravity field at the earth's surface by including terrestrial observations or, e.g., satellite altimetry. The detailed gravity field at the earth's surface cannot be adequately described by a spherical-harmonic expansion, neither from a theoretical point of view --because the convergence cannot be guaranteed--nor from a practical point of view--because, if at all possible, such an expansion would require an excessively high number of terms, which is beyond the capacity of any present digital computer.

Hence, the general replacement of Hilbert space by a finite-dimensional space is neither theoretically nor practically feasible.

It is also not necessary since, as we have seen, the practical collocation formulas are finite-dimensional matrix formulas. The approach of (Moritz, 1972) works entirely with finite matrices; the Hilbert space character expresses itself only in the fact that covariances are propagated, not by matrix operations, but by linear operations (such as differentiation) of covariance functions. See also (Moritz, 1976, sec.4).

A feature of collocation that is theoretically most interesting is that, from a formal-geometrical point of view, least-squares collocation can be considered as a problem of least-squares adjustment <u>in Hilbert space</u> (Krarup, 1969, pp.34-41).

We also mention that the usual least-squares solution of overdetermined systems of linear equations can be considered as a "quasi-solution" according to the theory of improperly posed problems (Tikhonov and Arsenin, 1977, Chapter I).

Discrete versus continuous data. In general, geodetic data are discrete measurements of a finite number; the present report is based on this situation. There are exceptions, for instance, continuous recordings of data profiles; they are usually converted into discrete data by a representative selection.

A practically more important case is, for instance, a very dense gravity survey around a station at which precise deflections of the vertical are to be computed. In such cases, a combination of collocation with other methods, such as integral formulas (Moritz, 1975), might be appropriate.

REFERENCES

Collatz,L. (1966) The Numerical Treatment of Differential Equations, Springer, Berlin.

Cordova, W.R. (1977) New concepts in geodetic instrumentation. In: The Changing World of Geodetic Science (U.A. Uotila, ed.), Report No.250, Dept. of Geodet. Sci., Ohio State Univ., vol.I, pp.73-97.

Dermanis, A. (1976) Probabilistic and deterministic aspects of linear estimation in geodesy. Report No.244, Dept. of Geodet. Sci., Ohio State Univ.

Dieudonné, J. (1960) Foundations of Modern Analysis. Academic Press, New York.

Eeg, J., and T. Krarup (1975) Integrated geodesy. Methoden und Verfahren der mathemat. Physik, 13, pp.77-123, Bibliographisches Institut, Mannheim.

Giacaglia, G.E.O., and C.A. Lundquist (1972) Sampling functions for geophysics. Smithson. Astrophys. Observatory, Spec. Report No.344.

Grafarend, E. (1972) Nichtlineare Prädiktion. Zeitschrift für Vermessungswesen, 97, pp.245-255.

Grafarend, E. (1977) Space-time differential geodesy. In: The Changing World of Geodetic Science (U.A. Uotila, ed.), Report No.250, Dept. of Geodetic Sci., Ohio State Univ., vol.I, pp.150-216.

Grafarend, E. (1978) Operational geodesy. In (Moritz and Sünkel, 1978), pp.235-284.

Hardy, R.L. (1976) Geodetic applications of multiquadric equations. Report, Engineering Res. Inst., Iowa State Univ., Ames.

Heiskanen, W.A., and H. Moritz (1967) Physical Geodesy. Freeman, San Francisco.

Kantorovich, L.V., and G.P. Akilov (1964) Functional Analysis in Normed Spaces. Macmillan, New York.

Kaula, W.M. (1966) Global harmonic and statistical analysis of gravimetry. Gravity Anomalies: Unsurveyed Areas, Geophys. Monograph No.9, Am. Geophys. Union, Washington, pp.58-67.

Kaula, W.M. (1967) Theory of Satellite Geodesy, Blaisdell, Waltham, Mass..

Krarup, T. (1968) A framework for least-squares determination of the potential of the earth. Report to the members of I.A.G. Special Study Group 5.31, unpublished.

Krarup, T. (1969) A contribution to the mathematical foundation of physical geodesy. Publ.No.44, Danish Geodetic Inst., Copenhagen.

Krarup, T. (1978) Some remarks about collocation. In (Moritz and Sünkel, 1978), pp.193-209.

Kryński, J. (1978) Possibilities of low-low satellite tracking for local geoid improvement. Mitteilungen der geodät. Institute der Techn. Univ. Graz, No.31.

Lavrentiev, M.M. (1967) Some Improperly Posed Problems of Mathematical Physics. Springer, Berlin.

Loomis, L.H., and S. Sternberg (1968) Advanced Calculus. Addison-Wesley, Reading, Mass..

Macdoran, P.F. (1973) Very long baseline interferometry (VLBI) applications to secular geodynamics and earth strain measurement. Proc. Symposium on Earth's Gravitational Field and Secular Variations in Position, pp.380-395, School of Surveying, Univ. of New South Wales, Sydney, Australia.

Meissl, P. (1975) Elements of functional analysis. Methoden und Verfahren der mathemat. Physik, 12, pp.19-78, Bibliographisches Institut, Mannheim.

Meschkowski, H. (1962) Hilbertsche Räume mit Kernfunktion. Springer, Berlin.

Moritz, H. (1966) Zur geometrischen Deutung der Minimumsprinzipien der Ausgleichungsrechnung. Zeitschrift für Vermessungswesen, 91, pp.293-296.

Moritz, H. (1972) Advanced least-squares methods. Report No.175, Dept. of Geodet. Sci., Ohio State Univ..

Moritz, H. (1975) Integral formulas and collocation. Ibid., Report No.234.

Moritz, H. (1976) Least-squares collocation as a gravitational inverse problem. Ibid., Report, No.249.

Moritz, H. (1978a) Statistical foundations of collocation. Ibid., Report No.272.

Moritz, H. (1978b) Introduction to interpolation and approximation. In (Moritz and Sünkel, 1978), pp.1-45.

Moritz, H. (1978c) Least-squares collocation. Reviews of Geophysics.

Moritz, H., and F. Sansò (1978) Dialogue on collocation. Proceedings, Symposium on Mathematical Geodesy, Assisi, June 1978.

Moritz, H., and H. Sünkel, eds. (1978) Approximation Methods in Geodesy. Herbert Wichmann Verlag, Karlsruhe.

Nashed, Z. (1974) Approximate regularized solutions to improperly posed linear integral and operator equations. In: Constructive and Computational Methods for Differential and Integral Equations (D.L. Colton and R.P. Gilbert, eds.), Lecture Notes in Mathematics, vol.430, Springer, Berlin.

Neyman, Yu.M. (1977) A variational method for solving discrete problems of physical geodesy (in Russ.), Geodeziya i Aerofotosyemka, pp.21-27.

Sansò, F. (1978) The minimum mean square estimation error principle in physical geodesy (stochastic and nonstochastic interpretation). Proceedings, Symposium on Mathematical Geodesy, Assisi, June 1978.

Schwarz, K.P. (1978) Geodetic improperly posed problems and their regularization. Lecture Notes, International School of Advanced Geodesy, Erice (Sicily), May-June 1978.

Sünkel, H. (1978) Approximation of covariance functions by non-positive-definite functions. Report No.271, Dept. of Geodetic Sci., Ohio State Univ.

Tikhonov, A.N., and V.Y. Arsenin (1977) Solutions of Ill-Posed Problems. Wiley, New York.

Tscherning, C.C. (1975) Application of collocation: determination of a local approximation to the anomalous potential using "exact" astro-gravimetric collocation. Methoden und Verfahren der mathemat. Physik, 14, pp.83-110, Bibliographisches Institut, Mannheim.

Tscherning, C.C. (1977) A note on the choice of norm when using collocation for the computation of approximations to the anomalous potential. Bull. Géod., 51, pp.137-147.

Tscherning, C.C. (1978) Introduction to functional analysis with a view to its applications in approximation theory. In (Moritz and Sünkel, 1978), pp.157-191.